CD METROLOGY METHOD

Inventors: Kenneth Johnson

Priority Claim

5   The present application claims priority to U.S. Provisional Patent Application Serial No. 60/443,028, filed January 28, 2003, the disclosure of which is incorporated in this document by reference.

Technical Field

           The subject invention relates to systems used to inspect and analyze

10   semiconductor wafers. In particular, the present invention relates to methods for rapidly evaluating scatterometry data to perform critical dimension and other measurements.

Background of the Invention

           As semiconductor geometries continue to shrink, manufacturers have increasingly turned to optical techniques to perform non-destructive inspection and analysis of

15   semiconductor wafers. Techniques of this type, known generally as optical metrology, operate by illuminating a sample with an incident field (typically referred to as a probe beam) and then detecting and analyzing the reflected energy. Ellipsometry and reflectometry are two examples of commonly used optical techniques. For the specific case of ellipsometry, changes in the polarization state of the probe beam are analyzed.

20   Reflectometry is similar, except that changes in intensity are analyzed. Ellipsometry and reflectometry are effective methods for measuring a wide range of attributes including information about thickness, crystallinity, composition and refractive index. The structural details of ellipsometers are more fully described in U.S. Patent Nos. 5,910,842 and 5,798,837 both of which are incorporated in this document by reference.

25   As shown in Figure 1, a typical ellipsometer or reflectometer includes an illumination source that creates a mono or polychromatic probe beam. The probe beam is focused by one or more lenses to create an illumination spot on the surface of the sample under test. A second lens (or lenses) images the illumination spot (or a portion of the illumination spot) to a detector. The detector captures (or otherwise processes) the

30   received image. A processor analyzes the data collected by the detector. The structural

TWI-33010

details of ellipsometers are described more fully in U.S. Patent No. 5,910,842 and 5,798,837 both of which are incorporated in this document by reference.

Scatterometry is a specific type of optical metrology that is used when the structural geometry of a sample creates diffraction (optical scattering) of the incoming probe beam. Scatterometry systems analyze diffraction to deduce details of the structures that cause the diffraction to occur. Various optical techniques have been used to perform optical scatterometry. These include broadband spectroscopy (U.S. Patent Nos. 5,607,800; 5,867,276 and 5,963,329), spectral ellipsometry (U.S. Patent No. 5,739,909) single-wavelength optical scattering (U.S. Patent No. 5,889,593), and spectral and single-wavelength beam profile reflectance and beam profile ellipsometry (U.S. Patent No. 6,429,943). Scatterometry, in these cases generally refers to optical responses in the form of diffraction orders produced by period structures, that is, gratings on the wafer. In addition it may be possible to employ any of these measurement technologies, e.g., single-wavelength laser BPR or BPE, to obtain critical dimension (CD) measurements on non periodic structures, such as isolated lines or isolated vias and mesas. The above cited patents and patent applications, along with PCT Application WO 03/009063, US Application 2002/0158193, US Application 2003/0147086, US Application 2001/0051856 Al, and PCT Application WO 01/97280 are all incorporated herein by reference.

Most scatterometry systems use a modeling approach to transform scatterometry results into critical dimension measurements. For this type of approach, a theoretical model is defined for each physical structure that will be analyzed. The theoretical model predicts the empirical measurements (scatterometry signals) that scatterometry systems would record for the structure. A rigorous coupled wave theory can be used for this calculation. The theoretical results of this calculation are then compared to the measured data (typically in normalized form). To the extent the results do not match, the theoretical model is modified and the theoretical data is calculated once again and compared to the empirical measurements. This process is repeated iteratively until the correspondence between the calculated theoretical data and the empirical measurements reaches an acceptable level of fitness. At this point, the characteristics of the theoretical model and the physical structure should be very similar.

TWI-33010

Evaluation of the theoretical models is a complex task, even for relatively simple structures. As the models become more complex (particularly as the profiles of the walls of the features become more complex) the calculations can become extremely time consuming. Even with high-speed processors, real-time evaluation of these calculations

5   can be difficult. Analysis on a real-time basis is very desirable so that manufacturers can immediately determine when a process is not operating correctly. The need is becoming more acute as the industry moves towards integrated metrology solutions wherein the metrology hardware is integrated directly with the process hardware.

A number of approaches have been developed to overcome the calculation

10   bottleneck associated with the analysis of scatterometry results. Many of these approaches have involved techniques for improving calculation throughput, such as parallel processing techniques. An approach of this type is described in a co-pending PCT application WO 03/009063, (incorporated herein by reference) which describes distribution of scatterometry calculations among a group of parallel processors. In the

15   preferred embodiment, the processor configuration includes a master processor and a plurality of slave processors. The master processor handles the control and the comparison functions. The calculation of the response of the theoretical sample to the interaction with the optical probe radiation is distributed by the master processor to itself and the slave processors.

20   For example, where the data is taken as a function of wavelength, the calculations are distributed as a function of wavelength. Thus, a first slave processor will use Maxwell's equations to determine the expected intensity of light at selected wavelengths scattered from a given theoretical model. The other slave processors will carry out the same calculations at different wavelengths. Assuming there are five processors (one

25   master and four slaves) and fifty wavelengths, each processor will perform ten such calculations per iteration.

Once the calculations are complete, the master processor performs the best fit comparison between each of the calculated intensities and the measured normalized intensities. Based on this fit, the master processor will modify the parameters of the

30   model as discussed above (changing the widths or layer thickness). The master processor

TWI-33010

will then distribute the calculations for the modified model to the slave processors. This sequence is repeated until a good fit is achieved.

This distributed processing approach can also be used with multiple angle of incidence information. In this situation, the calculations at each of the different angles of incidence can be distributed to the slave processor. Techniques of this type are an effective method for reducing the time required for scatterometry calculations. At the same time, the speedup provided by parallel processing is strictly dependent on the availability (and associated cost) of multiple processors. Amdahl's law also limits the amount of speedup available by parallel processing since serial portions of the program are not improved. At the present time, neither cost nor ultimate speed improvement is a serious limitation for parallel processing techniques. As the complexity of the geometry increases, however it becomes increasingly possible that computational complexity will outstrip the use of parallel techniques alone.

Another approach is to use pre-computed libraries of predicted measurements. This type of approach is discussed in U. S. Patent No. 6,483,580, (incorporated by reference) as well as the references cited therein. In this approach, the theoretical model is parameterized to allow the characteristics of the physical structure to be varied. The parameters are varied over a predetermined range and the theoretical result for each variation to the physical structure is calculated to define a library of solutions. When the empirical measurements are obtained, the library is searched to find the best fit.

The use of libraries speeds the analysis process by allowing theoretical results to be computed once and reused many times. At the same time, library use does not completely solve the calculation bottleneck. Construction of libraries is time consuming, requiring repeated evaluation of the same time consuming theoretical models. Process changes and other variables may require periodic library modification or replacement at the cost of still more calculations. For these reasons, libraries are expensive (in computational terms) to build and to maintain. Libraries are also necessarily limited in their resolution and can contain only a finite number of theoretical results. As a result, there are many cases where empirical measurements do not have exact library matches.

One approach for dealing with this problem is to generate additional theoretical results in real-time to augment the theoretical results already present in the library. PCT

TWI-33010

WO 02/27288, published April 4, 2002 suggests a combined approach of this type where a coarse library is augmented by real-time regression. Combined approaches have typically improved accuracy, but slowed the scatterometry process as theoretical models are evaluated in real-time.

5   Another approach is to use a library as a starting point and generate missing results as needed. For example, U.S. Patent No. 5,867,276 describes a system of training a library to permit linear estimations of solutions. Alternatively, the use of interpolation is described in U.S. Patent Application 2002/0038196, published March 28, 2002. The use of interpolation avoids the penalty associated with generating results in real-time, but

10   may sacrifice accuracy during the interpolation process. The latter publications are incorporated herein by reference.

For these reasons and others, there is a continuing need for faster methods for computing results for scatterometry systems. The need for faster evaluation methods will almost certainly increase as models become more detailed to more accurately reflect

15   physical structures.

Summary of the Invention

The present invention provides a method for rapidly evaluating data gathered during scatterometry and related optical metrology procedures. For this method, a theoretical model is used to construct a database for each sample that is to be evaluated.

20   The theoretical model predicts the output (reflected) electromagnetic field that is generated when an incident field is applied to the sample. The relationship between these two fields is referred to in this document as the sample's optical response. The theoretical model is parameterized and each parameter corresponds to a characteristic of the sample such as line width, line profile or layer thickness. The database is a mapping between a

25   series of predetermined parameter sets and their associated optical responses. It is constructed by repeatedly evaluating the theoretical model for each of the predetermined parameter sets and storing the resulting optical responses in association with their corresponding parameter sets.

In general, the database is constructed in an offline, non-realtime basis that

30   precedes the actual measurement of the sample. At the time of measurement, the optical metrology system measures the optical response of the sample. The database is then

TWI-33010

searched to locate the theoretical optical response and associated parameter set that most closely matches the measured optical response.

An interpolation module is used to refine the results obtained from the database. This refinement is useful because the optical response as measured does not necessarily

5   correspond to any optical response included in the database. To reduce this discrepancy, the interpolation module constructs a series of trial parameter sets. Using an interpolation model, it then constructs an interpolated optical response for each trial parameter set. The process is repeated until an interpolated optical response is computed that closely matches the measured optical response. The calculation that translates each trial

10   parameter set into its associated interpolated optical response ("interpolative refinement") relies only on the database-stored optical responses, without reference to the sample's underlying theoretical model. In this way, the computational complexity of rigorous analysis avoided during the interpolation phase. (The database interpolation method is detailed in U.S. Patent Application 2002/0038196).

15   A theoretical refinement module is used to refine the results obtained from the interpolation module. This refinement is useful because the optical response generated by the interpolation module may contain errors generated by the interpolation process. As a result, the interpolated optical response does not necessarily correspond to the exact optical response that would be generated by the particular sample being analyzed. To

20   reduce this discrepancy, the theoretical refinement module uses a theoretical model of the sample. Typically, this is the same model that is used to construct the database but it is also possible to use a model that is optimized for runtime performance. The model is repeatedly evaluated using a series of associated parameter sets. Each evaluation produces an optical response that is compared to the measured optical response. This

25   process continues until a desired goodness of fit is achieved. (The refinement may be performed according to the method of PCT application WO 03/009063.)

The parameter set produced by the theoretical refinement module closely matches the actual measurement parameters of the sample. In practice, this level of accuracy may not always be required. For this reason, it is generally desirable to configure the data

30   evaluation method to operate adaptively and use any combination of database search,

interpolative refinement and theoretical refinement to produce results having any desired accuracy.

In this way, the present invention avoids the inaccuracies associated with traditional interpolation-based analysis and without incurring the computational complexity associated with real-time database supplementation.

Brief Description of the Drawings

Figure 1 shows a prior art optical metrology system.

Figure 2 shows a generic sample suitable for analysis using the present invention.

Figure 3 is a flow diagram showing the modules used by the data analysis of the present invention.

Figure 4 is a block diagram showing the database searching module used by the present invention.

Figure 5 is a block diagram showing the interpolative refinement module used by the present invention.

Figure 6 is a block diagram showing the theoretical refinement module used by the present invention.

Detailed Description of the Preferred Embodiments

An aspect of the present invention provides a method for rapidly evaluating data gathered during scatterometry and related optical metrology methods. Figure 2 shows a generic sample 200 of the type typically analyzed by scatterometry systems. As shown in Figure 2, sample 200 includes a surface structure 202. For this example, surface structure 202 is a grating that is periodic in the X direction and uniform (exhibits translational symmetry) in the Y direction. In general, the surface structure may be composed of any combination of periodic or isolated, two or three-dimensional structures. Sample 200 is covered by an incident medium (not shown) that is typically air but may be vacuum, gas, liquid, or solid (such as an overlaying layer or layers). Below surface structure 202, sample 200 may include one or more layers constructed using one or more different materials. In Figure 2, the internal layers are labeled 204a through 204c. At the bottom of internal layers 204, sample 200 includes a final layer, known as a substrate 206.

Optical metrology systems apply an incident field ψ_{in} to sample 200 and measure the resulting the output field ψ_{out}. By analyzing the output field ψ_{out} optical metrology systems are able to ascertain underlying measurement characteristics about sample 200, such as line sizes and spacings, film properties and thicknesses. The present invention

5   provides a method for rapidly correlating measurements of the output field ψ_{out} to characteristics of this type. Prior to using this method, a library or database is constructed using a parameterized model to represent sample 200. The parameterized model allows a theoretical optical response of sample 200 to be calculated as a function of a set of one or more parameters. Each parameter within the parameter set corresponds to a measurement

10   parameter associated with sample 200 (such as line width, line profile, layer thickness, etc.). The parameterized model is used to build a database. For this task, a series of parameter sets are defined. The theoretical model is evaluated for each parameter set and the resulting optical response is stored (in association with the corresponding parameter set) in the database. The finished database functions as a mapping between the series of

15   parameter sets (measurement characteristics) and their associated optical responses (measurements of the output field ψ_{out}).

After the database has been constructed, it becomes one of the runtime components used by the data evaluation method of the present invention. To illustrate, Figure 3 shows a representative implementation of the data evaluation method including

20   a database 302 (constructed as described), paired with modules for database searching 304, interpolated refinement 306 and theoretical refinement 308. Database searching module 304 performs searching or lookup on database 302. As shown in Figure 4, database searching module 304 is supplied with an empirical optical response as input. Typically, this consists of one or more empirical measurements obtained during

25   inspection of a sample.

In response, database searching module 304 uses a bet-fit strategy to locate the closest matching theoretical optical response within database 302. Database searching module 304 then returns the closest matching theoretical optical response along with its associated parameter set. In this way, database searching module 304 provides a

30   mechanism for mapping measured optical responses to matching parameters sets. The

mechanism operates with a degree of built-in granularity and only returns optimal results when a target optical response exactly matches a theoretical optical response that is included in database 302.

Interpolated refinement module 306 is used to enhance the accuracy of results

5   beyond what is available by searching database 302. Interpolated refinement module 306 includes two main components: 1) a fitting optimization algorithm, and 2) an interpolation model. As shown in Figure 5, the fitting optimization algorithm accepts a target optical response and an initial parameter set as inputs. During the most common use of the data evaluation method, the target optical response and the initial parameter set

10   are the inputs and outputs of database searching module 304 (i.e., the target optical response is obtained during inspection of the sample and the initial parameter set is the corresponding database entry). For other applications, the target optical response and initial parameter set may be user-supplied.

In response, the fitting optimization algorithm generates an interpolated parameter

15   set and an interpolated optical response as outputs. To perform this task, the fitting optimization algorithm starts with the initial parameter set and generates a succession of trial interpolation parameter sets. Each trial parameter set is passed, by the fitting optimization algorithm to the interpolation model for evaluation. The interpolation model calculates a corresponding interpolated optical response for each trial parameter set. The

20   optimization algorithm compares each interpolated optical response to its target optical response. The algorithm selects the trial parameter sets, based on a comparison error minimization method, to iteratively reduce a defined comparison error metric until a defined termination criterion is satisfied. The trial parameters used to achieve the desired goodness of fit (as well as the associated interpolated optical response), are returned as

25   the outputs of the fitting optimization algorithm (and interpolated refinement module 306).

As mentioned, the interpolation model returns an interpolated optical response for each trial parameter set supplied by the fitting optimization algorithm. To perform this task, the interpolation model performs interpolation (or, in some cases extrapolation)

30   using the parameter sets in database 302. This allows the interpolation model to construct interpolated optical responses for trial parameter sets, even for trial parameter sets that

are not included in database 302. The construction of interpolated optical responses is performed without reference to the underlying theoretical model. In this way, the interpolation model (and in turn, interpolated refinement module 306) avoids the computational overhead of direct theoretical modeling of sample 200. Various types of interpolation models may be used. In particular, the use of multi-linear or multi-cubic based models is appropriate. Methods of this type are more fully described in U.S. Patent Application 2002/0038196, and U.S. Patent Application No. 10/611,298 both of which are incorporated in this document by reference.

Theoretical refinement module 308 is used to enhance the accuracy of results beyond what is available by interpolated refinement module 306. As shown in Figure 6, theoretical refinement module 308 accepts a target optical response and an initial parameter set as inputs. During the most common use of the data evaluation method, the target optical response is the empirical optical response obtained during inspection of the sample and the initial parameter set is the output of interpolated refinement module 306. For other applications, the target optical response and initial parameter set may be user-supplied. The initial parameter set may also be the output of database searching module 304.

Using these inputs, theoretical refinement module 308 constructs a theoretical parameter set and a theoretical optical response as outputs. To perform this computation, theoretical refinement module 308 uses a theoretical model of sample 200. In most cases, the theoretical model is the same as the parameterized model used to construct database 302. There may, however, be implementations where different models are used for these two applications. As previously described, the theoretical model is parameterized allowing its modeled characteristics to be systematically altered. Using the initial parameter set as a baseline, theoretical refinement module 308 repeatedly evaluates the theoretical model with varying parameters. Each evaluation creates a theoretical optical response. With each repetition, the theoretical refinement module 308 compares this theoretical optical response to the target optical response. The process continues until a desired goodness of fit has been achieved. The parameters used to achieve the desired goodness of fit (as well as the associated theoretical optical response) are returned as the outputs of the fitting optimization algorithm. The theoretical refinement may be

performed according to the method of PCT application WO 03/009063, which is incorporated in this document by reference.

The combination of database 302, database searching module 304, interpolated refinement module 306 and theoretical refinement module 308, provide a flexible

5   approach for mapping optical responses to associated measurement parameters. As shown in Figure 3, a typical application of these components begins with empirical measurements (designated 310 in Figure 3). Typically, measurements of this type would include optical response information gathered by an optical metrology tool such as a ellipsometer or reflectometer. The empirical measurements would typically be evaluated

10   by initial processing by database searching module 304, followed by refining by interpolated refinement module 306 and theoretical refinement module 308 to produce a set of measurements parameters (designated 312 in Figure 3). Thus, the typical sequence is empirical measurement followed by database lookup, interpolative refinement and theoretical refinement. Depending on the particular level of accuracy required as well as

15   knowledge about the particular sample 200 involved, any combination of these processes may be bypassed. The particular choice of processes is intended to be dynamically configurable to suit a range of analysis needs. Alternately, the same set of components (in any combination) may be used to analyze user-supplied measurements (e.g., parameter estimates (designated 314 in Figure 3)).

20   In this way, the present invention avoids the inaccuracies associated with traditional interpolation-based analysis and without incurring the computational complexity associated with real-time database supplementation.

It should also be noted that the combination of database 302, database searching module 304, interpolated refinement module 306 and theoretical refinement module 308

25   provide a synergistic benefit that exceeds the simple sum of these separate components. This is because theoretical refinement module 308 reduces the resolution required for database 302. In other words, database 302 may include fewer points sampled at coarser intervals. The use of interpolation, on the other hand, reduces the range of simulations performed by theoretical refinement module 308. This results because the starting value

30   passed to theoretical refinement module 308 is generally closer to the desired result than would be possible without the use of interpolation.